Módulo de Estadística

Tema 8: Introducción a los contrastes de hipótesis

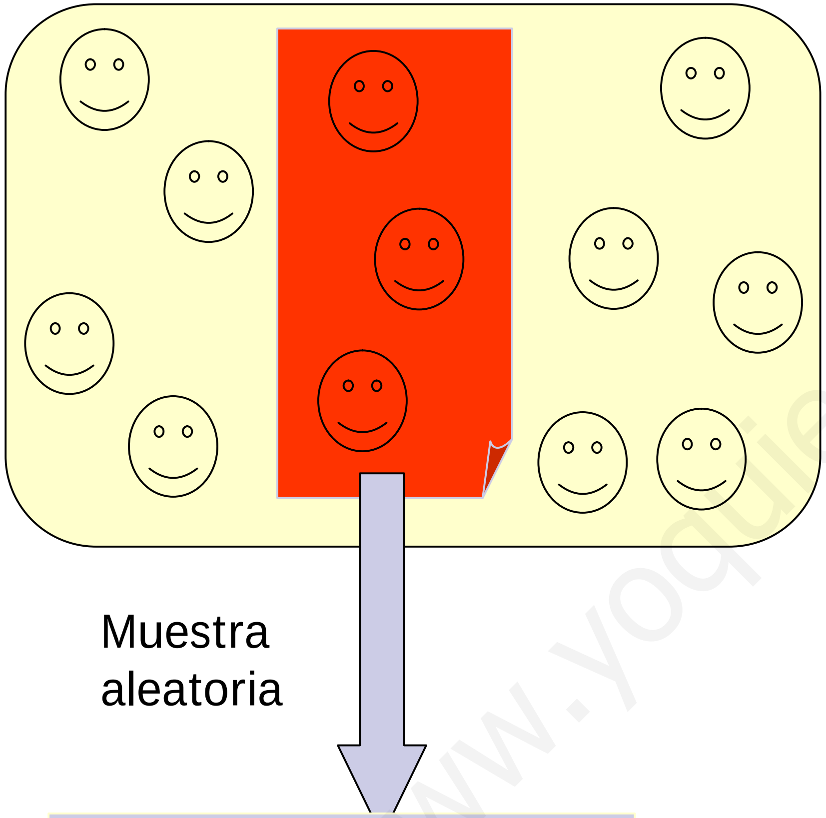
Objetivos del tema

- Introducir el concepto de contraste de hipótesis
- Diferenciar entre hipótesis nula y alternativa
- Nivel de significación
- Errores tipo I y errores tipo II

Contrastando una hipótesis

Son demasiados...

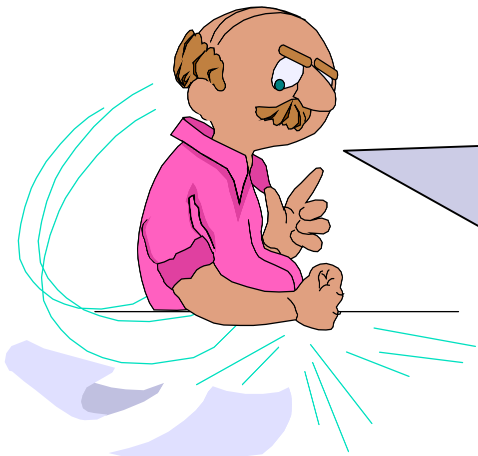
Creo que la edad media es 40 años...



Muestra aleatoria

¡Gran diferencia!

Rechazo la hipótesis



$$\overline{X} = 20 \text{ años}$$

- ## Contraste de Hipótesis:
  Procedimiento estadístico mediante el cual se investiga la aceptación o rechazo de una afirmación acerca de una o varias características de una población
- ## ¿Qué es una hipótesis?
  Una creencia o afirmación sobre la población, principalmente sus parámetros:
  - Media
  - Varianza
  - Proporción/Tasa

Identificación de hipótesis

- Hipótesis nula H_o
  - La que contrastamos
  - Los datos pueden refutarla
  - No debería ser rechazada sin una buena razón.
- Hip. Alternativa H_1
  - Niega a H_0
  - Los datos pueden mostrar evidencia a favor
  - No debería ser aceptada sin una gran evidencia a favor.

$$\begin{cases} H_0: & p = 50\% \\ H_A: & p \neq 50\% \end{cases}$$

$= \ , \ \leq \ , \ \geq$

$\neq \ , \ < \ , \ >$

¿Quién es H_0?

- Problema: ¿En un estudio realizado para determinar el estado de salud de una comunidad se entrevistó a 82 personas, preguntándoles acerca de su actividad física habitual. De las 82 personas encuestadas, 36 de ellas declararon practicar algún deporte de forma regular ?

  Contraste la hipótesis de si la proporción poblacional de práctica de deporte de forma habitual puede ser diferente de 0,50

- Solución:
  - Traducir a lenguaje estadístico: $p = 50\%$
  - Establecer su opuesto: $p \neq 50\%$
  - Seleccionar la hipótesis nula $H_0 : p = 50\%$

¿Quién es H_0?

- Problema: ¿El colesterol medio para la dieta mediterránea es 6 mmol/l, la media es diferente de 6 mmol/l?

- Solución:
  - Traducir a lenguaje estadístico: $\mu = 6$
  - Establecer su opuesto: $\mu \neq 6$
  - Seleccionar la hipótesis nula

$$H_0 : \mu = 6$$

Test de hipótesis o Estadístico del contraste (T)

Es una función de la m.a.s cuya distribución se utiliza para el contraste, y que es una función del estimador asociado al parámetro que se quiere contrastar

- El contraste concluirá formulando una regla de actuación
- 1) Se calcula el EC

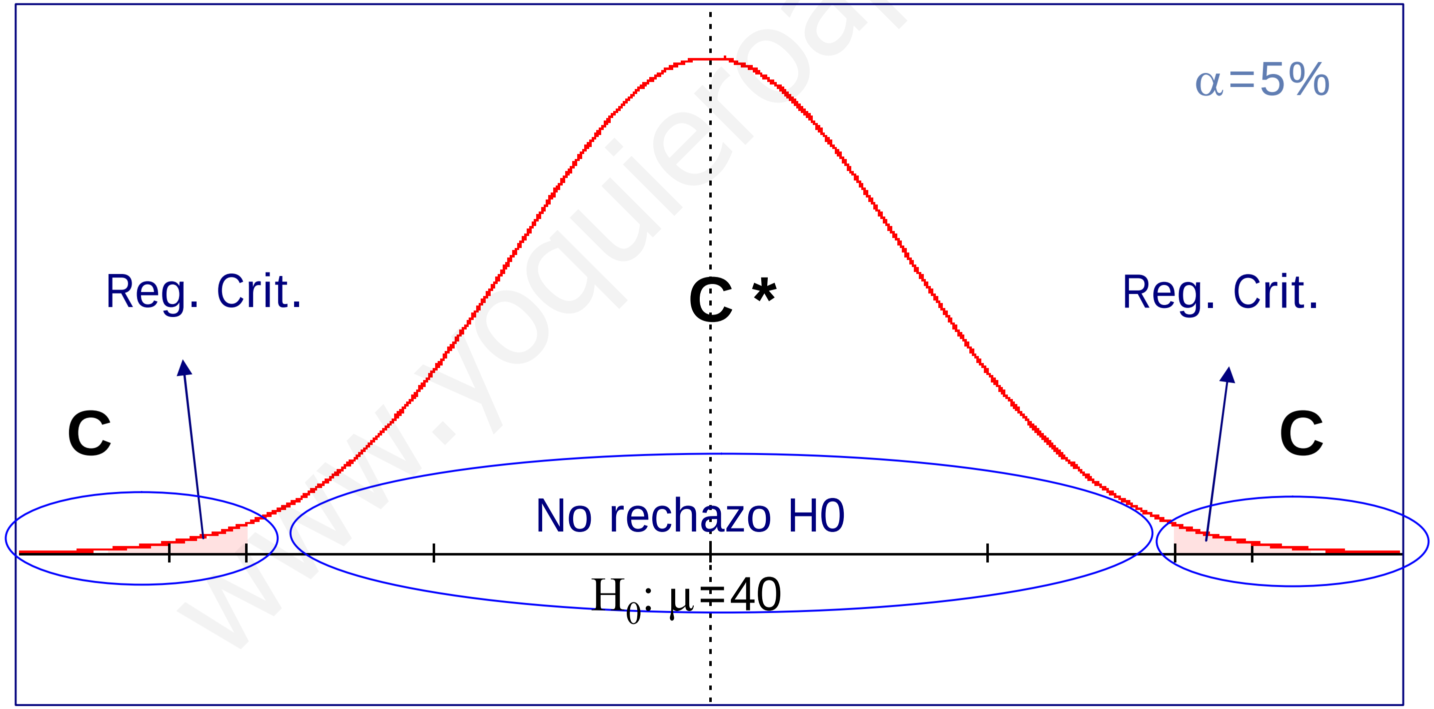
$$\begin{cases} \text{Acepta H0 si } T \in C^* \\ \text{Rechazar H0 si } T \in C \end{cases}$$

Región crítica y nivel de significación

Región crítica

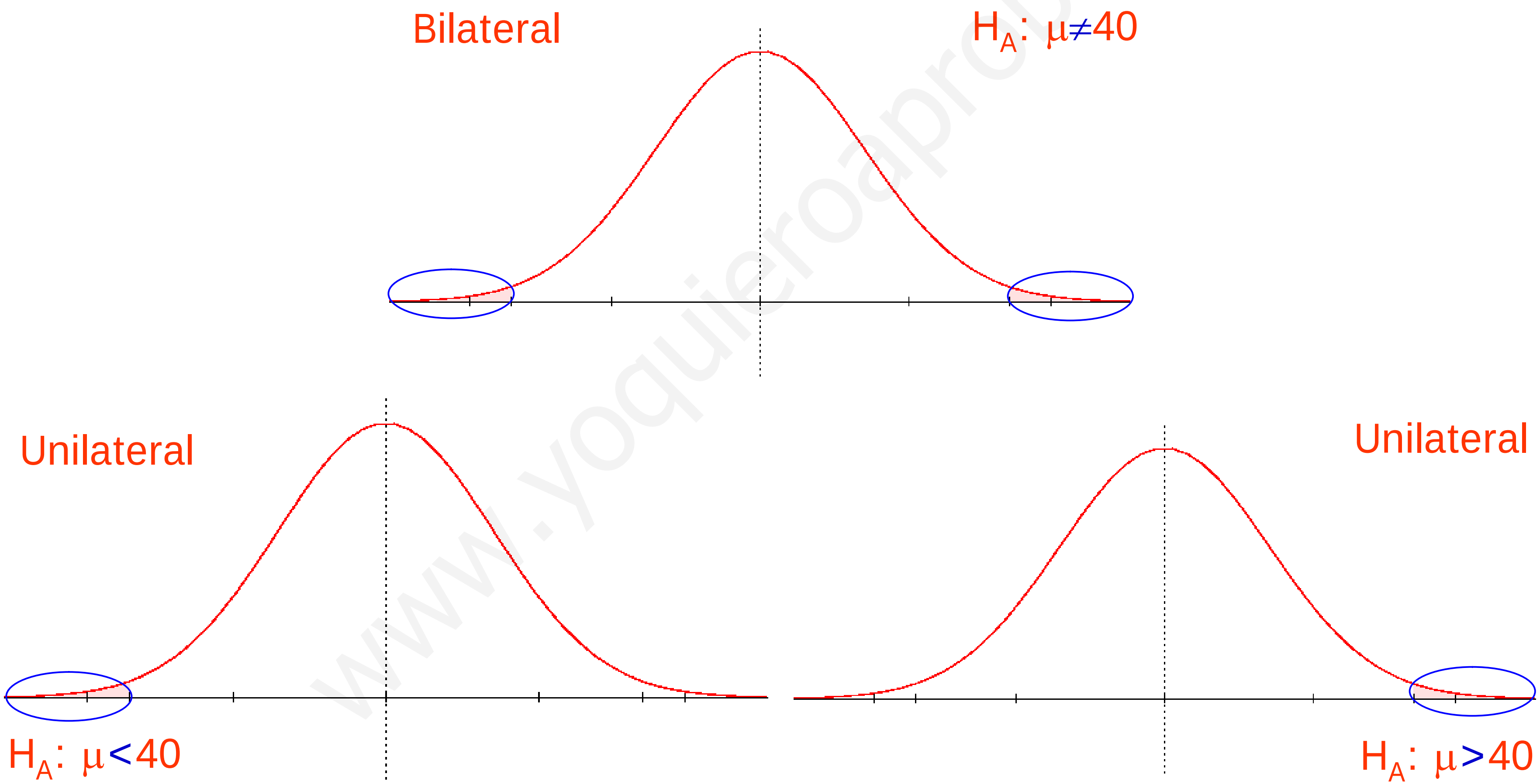
Nivel de significación: α

- Número pequeño: 1% , 5%
- Fijado de antemano por el investigador
- Es la probabilidad de rechazar H_0 cuando es cierta



Contrastes: unilateral y bilateral

La posición de la región crítica depende de la hipótesis alternativa

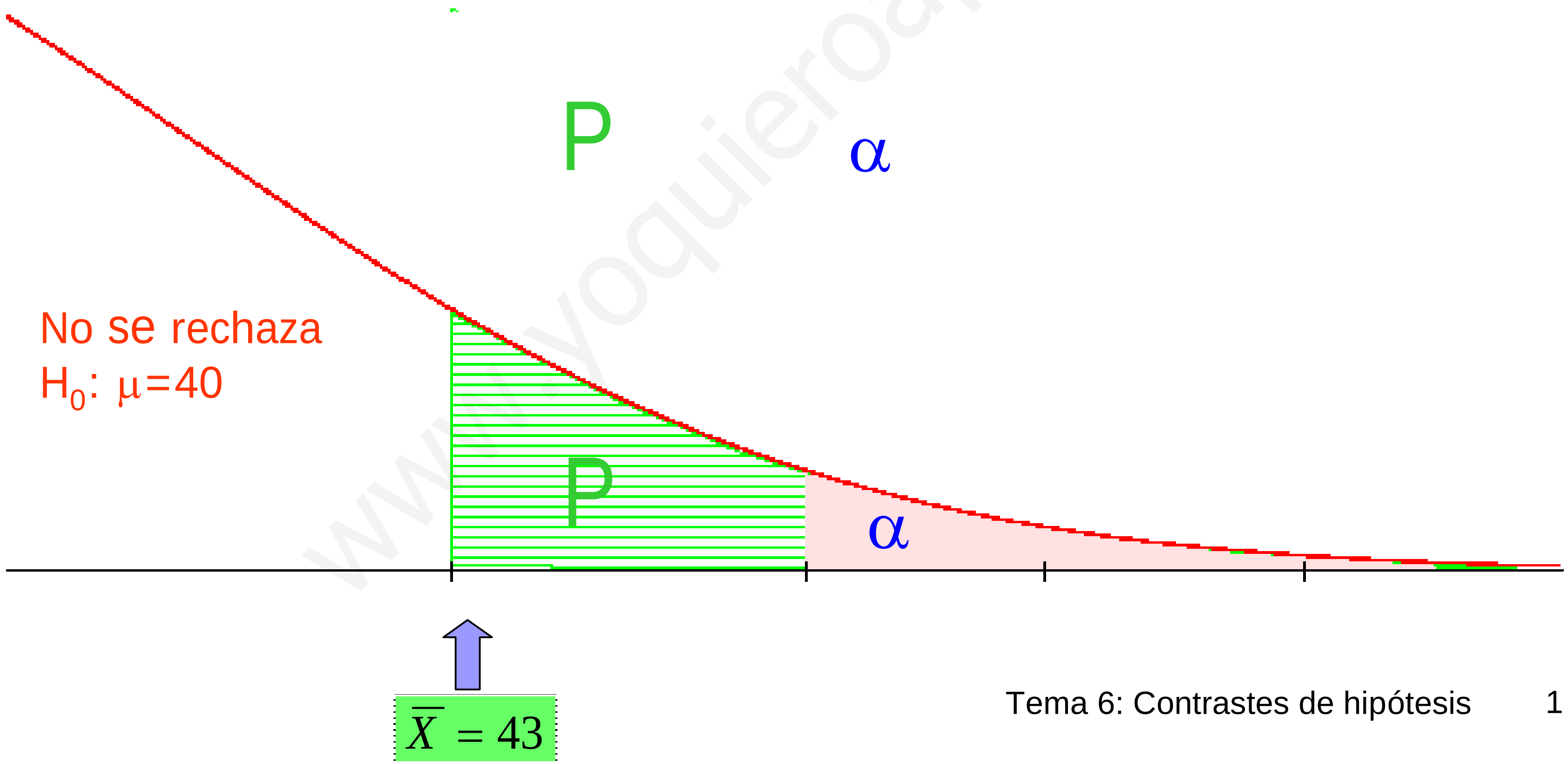


Significación: p

Probabilidad que el estadístico de contraste tome valores superiores al valor obtenido

p es conocido después de realizar el experimento aleatorio

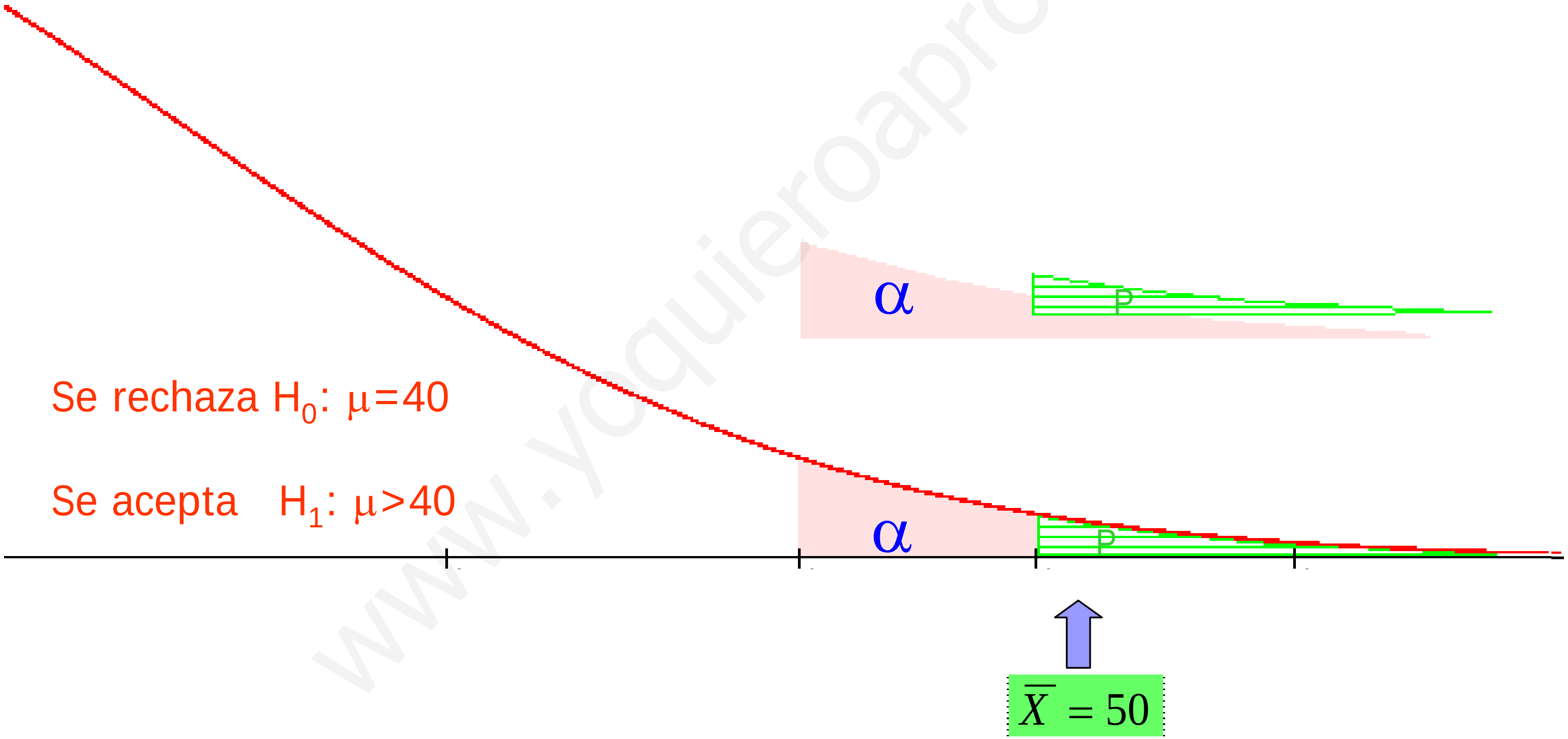
El contraste es no significativo cuando $p>\alpha$



Significación : p

El contraste es estadísticamente significativo cuando $p<\alpha$

Es decir, si el resultado experimental discrepa más de "lo tolerado" a priori.



Resumen: α, p y criterio de rechazo

- Sobre α
  - Es número pequeño, preelegido al diseñar el experimento
  - Conocido α sabemos todo sobre la región crítica
- Sobre p
  - Es conocido tras realizar el experimento
  - Conocido p sabemos todo sobre el resultado del experimento
- Sobre el criterio de rechazo
  - Contraste significativo = p menor que α

Resumen: α, p y criterio de rechazo

as well as the Pearson one corrected for continuity.

A signification level of $P < 0{,}05$ was selected using tl
for Windows V-5, 1992 in a IBM compatible DX 486

MATERIAL Y METODO

Se realizó un estudio de tipo prospectivo y longitudir
estuvo constituido por todos los recién nacidos de m
servicio de Neonatología del Hospital Gineco-Obstétr
Ciudad de La Habana en el período comprendido ent
1993, a los que se le realizó un seguimiento en la se
Hospital por un equipo multitudinario.

La muestra del presente estudio estuvo constituida p
que cumplieran como criterio de inclusión haber com
primeros dos años de edad corregida como mínimo.

Con los 86 pacientes se conformaron 2 grupos:

a. Ventilados (26 pacientes): formado por aquello
mecánica a presión positiva intermitente durar
neonatal.

b. No Ventilados (60 pacientes): Integrado por lo
no recibieron ninguna modalidad de apoyo ver

- Sobre el criterio de rechazo
  - Contraste significativo = p menor que α

Tipos de error al tomar una decisión

		Realidad	
		Inocente	Culpable
veredicto	Inocente	OK	Error Menos grave
	Culpable	Error Muy grave	OK

Tipos de error al contrastar hipótesis

	Realidad	
	H_0 cierta	H_0 Falsa
No Rechazo H_0	**Correcto** El tratamiento no tiene efecto y así se decide.	**Error de tipo II** El tratamiento si tiene efecto pero no lo percibimos. Probabilidad β
Rechazo H0 Acepto H_1	**Error de tipo I** El tratamiento no tiene efecto pero se decide que sí. Probabilidad α	**Correcto** El tratamiento tiene efecto y el experimento lo confirma.

Conclusiones

- Las hipótesis no se plantean después de observar los datos.
- En ciencia, las hipótesis nula y alternativa no tienen el mismo papel:
  - H_0 : Hipótesis científicamente más simple.
  - H_1 :
- α debe ser pequeño
- Rechazar una hipótesis consiste en observar si $p<\alpha$
- Rechazar una hipótesis no prueba que sea falsa. Podemos cometer error de tipo I
- No rechazar una hipótesis no prueba que sea cierta. Podemos cometer error de tipo II
- Si decidimos rechazar una hipótesis debemos mostrar la probabilidad de equivocarnos.

Ejemplos

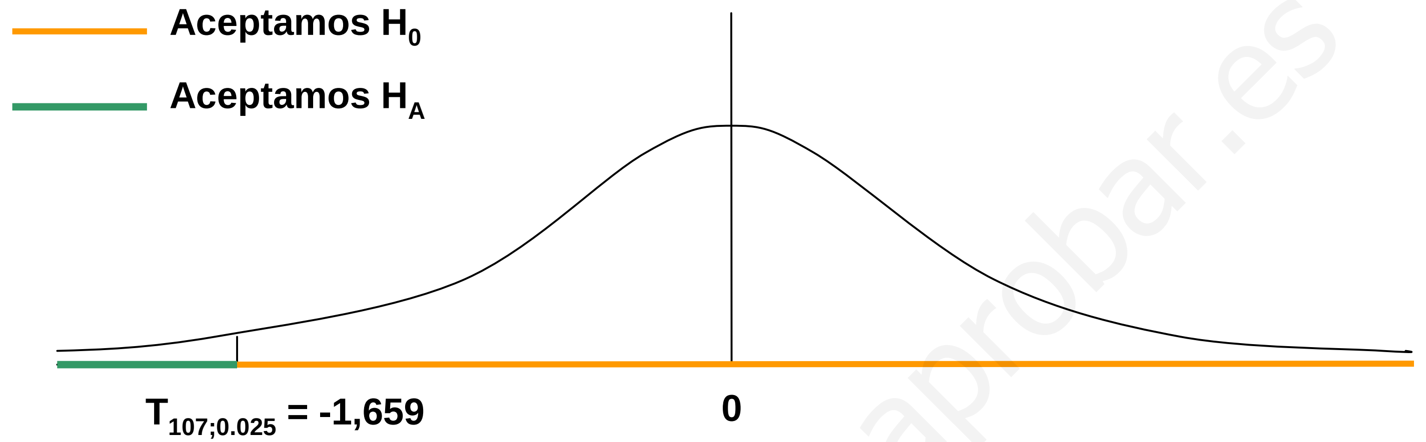
4.8. En un estudio sobre litiasis biliar realizado en Gandia y Real de Gandia se obtuvo información sobre 200 individuos. Sabiendo que de los 200 individuos 92 son hombres y 108 mujeres, y sabiendo que el nivel promedio de colesterol entre las 108 mujeres fue de 228.06 mg/100 ml con una desviación típica de 51.87

i) Sería admisible, a la vista del resultado obtenido, un valor promedio de colesterol superior a 235

1. Confeccionamos las hipótesis: $\begin{cases} \mathbf{H_o : \mu \leq 235} \\ \mathbf{H_A : \mu > 235} \end{cases}$

2. El percentil que buscamos sigue una distribución t-student con n-1 grados de libertad, al ser α=0,05 el percentil quebuscamos concretamente es:
   $t_{107;\,0,025}$ = -1,659

3. Estadístico de contraste $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{228{,}06 - 235}{51{,}87/\sqrt{108}} = -1{,}39$



Distribución de t suponiendo cierta H_0

Nuestro estadístico de contraste vale -1,39, por tanto aceptamos hipótesis nula de que la media poblacional es inferior a 235mg/100ml

Ejemplos

4.1. En un estudio realizado para determinar el estado de salud de una comunidad se entrevistó a 82 personas, preguntándoles acerca de su actividad física habitual. De las 82 personas encuestadas, 36 de ellas declararon practicar algún deporte de forma regular.

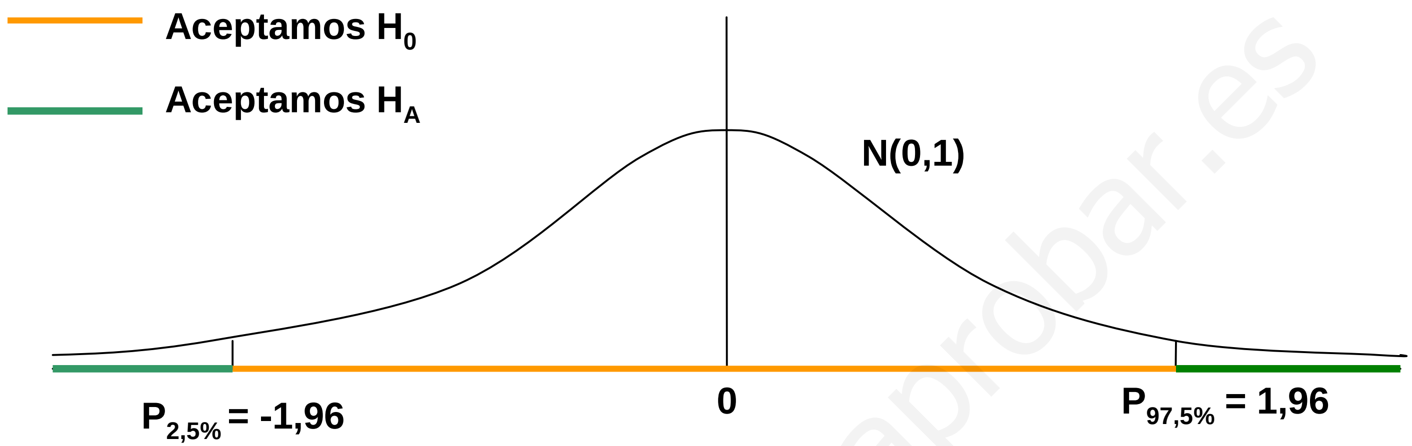
i) Contraste la hipótesis de si la proporción poblacional de práctica de deporte de forma habitual puede ser diferente de 0,50 con un nivel de confianza del 5 %.Calcule la *p* exacta para este contraste.

1. Confeccionamos las hipótesis: $\begin{cases} \mathbf{H_o : \rho = 0,50} \\ \mathbf{H_A : \rho \neq 0,50} \end{cases}$

2. Estimador puntal de la proporción: $\hat{p} = \dfrac{36}{82} = 0,439$

3. El percentil que buscamos sigue una distribución Normal con media 0 y varianza 1, al ser α=0,05 el percentil que buscamos concretamente es: = $\mathbf{Z_{97,5} = 1,96}$

4. Estadístico de contraste $$t = \frac{\hat{p} - p_o}{\sqrt{\dfrac{p_o(1-p_o)}{n}}} = \frac{0,439 - 0,5}{\sqrt{\dfrac{0,5(1-0,5)}{82}}} = \frac{-0,061}{0,055} = 1,1$$



Distribución de Z suponiendo cierta H_0

Nuestro estadístico de contraste vale 1,11, por tanto aceptamos hipótesis nula de que la proporción pueda ser 0,5.

Para el cálculo de la P-exacta hemos de obtener la probabilidad de que el estadístico de contraste Z tome valores superiores a 1,11 o inferiores a -1,11, por lo tanto:

P-exacta = P(Z>1,11) + (Z<-1,11) = 2x0,1335 = 0,1670

Ejemplos

En el estudio descrito en los ejercicios 4.1 y 4.2 se preguntó además por las horas de sueño de los encuestados. Los resultados expresados en media y desviación típica se recogen en la tabla adjunta y de forma separada para aquellos que declararon realizar ejercicio físico y para los que no:

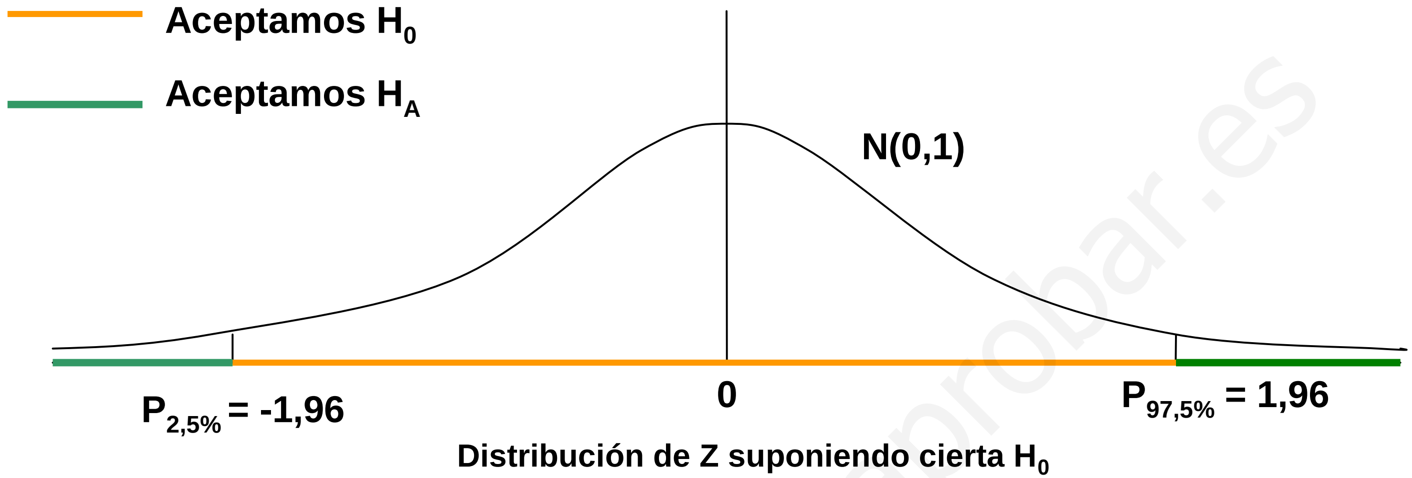
	Realizan ejercicio	**No realizan ejercicio**
Nº individuos	**36**	**46**
Media horas de sueño	**8.5 horas/día**	**7.2 horas/día**
D. Típica horas de sueño	**0.9 horas**	**0.8 horas**

i) A nivel de significación α=0.05, ¿existen diferencias significativas en los tiempos medios de sueño entre los individuos que realizan ejercicio físico y los que no?

1. Confeccionamos las hipótesis:

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

2. El percentil que buscamos sigue una distribución t-student con gl grados de libertad, al ser α=0.05 el percentil que buscamos concretamente es: $\mathbf{t_{gl;1-\alpha/2} = t_{73;0,975} = 1,992}$



3. Estadístico de contraste: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 6{,}81$

Nuestro estadístico de contraste vale 6,81, por tanto rechazamos hipótesis nula de que la medias poblacionales sean iguales

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2 / n_1\right)^2}{n_1} + \frac{\left(s_2^2 / n_2\right)^2}{n_2}} = 72{,}6 = 73$$

Ejemplos

4.2 Respecto de los datos del ejercicio 4.1, de las 82 personas encuestadas, 40 fueron hombres y el resto mujeres. De las 36 personas que declararon practicar ejercicio físico de forma regular, 10 eran mujeres y el resto hombres.

i) A nivel de significación 0.05, contraste la hipótesis de que la proporción de personas que realizan ejercicio físico de forma regular es mayor en hombres que en mujeres. Calcule la *p* exacta

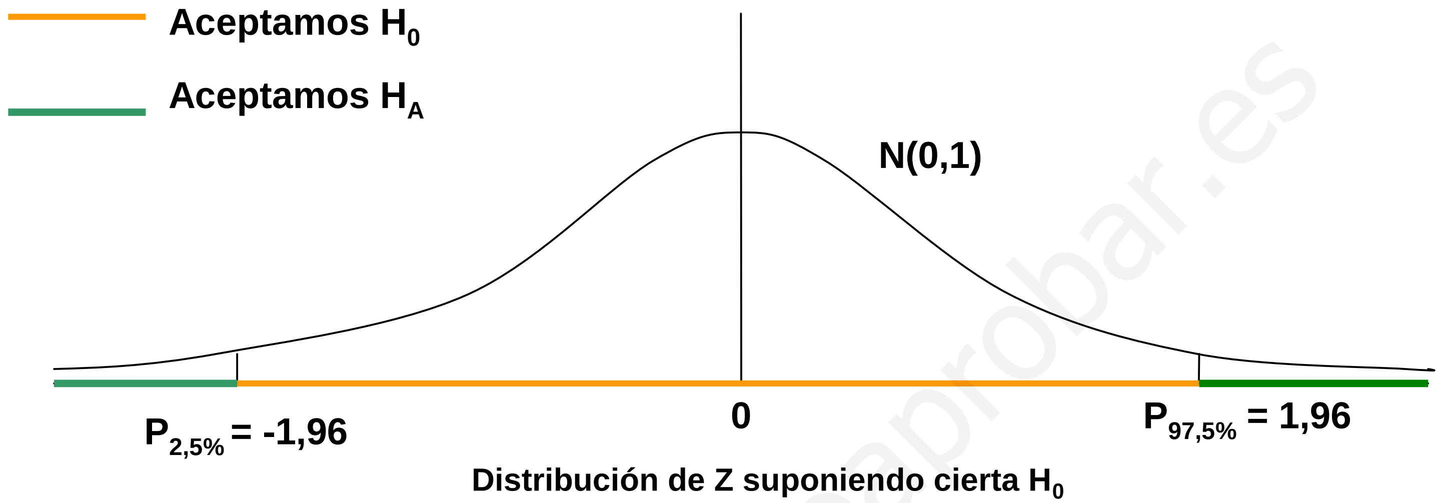
1. Confeccionamos las hipótesis: $\begin{cases} \mathbf{H_o : \rho_1 \leq \rho_2} \\ \mathbf{H_A : \rho_1 > \rho_2} \end{cases}$ $\begin{cases} \mathbf{H_o : \rho_1 - \rho_2 \leq 0} \\ \mathbf{H_A : \rho_1 - \rho_2 < 0} \end{cases}$

2. Estimadores puntuales de las proporción: $\hat{p}_1 = \frac{26}{40}$ $\hat{p}_2 = \frac{10}{42}$

3. El percentil que buscamos sigue una distribución Normal con media 0 y varianza 1, al ser α=0,05 el percentil que buscamos concretamente es: = $\mathbf{Z_{97,5} = 1,96}$

4. Estadístico de contraste

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} = \frac{0,65 - 0,24}{\sqrt{\frac{0,65(1-0,65)}{40} + \frac{0,24(1-0,24)}{42}}} = \frac{0,41}{\sqrt{0,0057 + 0,0043}} = \frac{0,41}{0,1} = 4,1$$



Nuestro estadístico de contraste vale 4,1, por tanto rechazamos hipótesis nula de que la proporciones sean iguales

Para el cálculo de la P-exacta hemos de obtener la probabilidad de que el estadístico de contraste Z tome valores superiores a 4,1 o inferiores a -4,1, por lo tanto:

P-exacta = P(Z>4,1) + (Z<-4,1) = 2x0,00003 = 0,00006